AB-44 - Paper

Immediate assessment of batch classification quality.

Francois J. LESCREVE *
Center for Recruitment and Selection
Belgian Armed Forces

1. Introductory section

Batch classification is used in selection settings where the data from a number of applicants are processed in order to decide which applicants will be assigned to a number of different vacant jobs. Batch classification, in opposition to sequential systems, processes the data of a whole group of applicants simultaneously. This is appropriate in settings where the enlistment is organized in groups, such as annual recruitments. Modern batch classification systems are generally composed of two major elements.

In the first element it is attempted to quantify the value of assigning a specific person to a specific job or a certain type of jobs. In the military, similar jobs are often labeled as *Military Occupation Specialties* (MOS) or as *trades*. The quantified values are called *payoff*-values and can be computed in several ways. Multiple linear regressions (MLR) are widely used. In MLR models, the payoffs usually are predicted performance scores on an external criterion that was used as dependent variable when designing the MLR model. Another method to produce payoff-values is the *Subject Matter Experts*-method (SME). In this method, subject matter experts are asked to give a specific weight to the selection variables for each MOS or trade. The payoffs can then be calculated as weighted sums. Artificial Neural Networks are also promising tools to generate payoff-values. The payoffs are computed for all person-job combinations and usually arranged in a payoff-matrix with the applicants as rows and the jobs as columns. The matrix is then squared by adding dummy-jobs.

When the payoff-matrix is ready, the second major element of the classification model is used. Since the matrix was squared it is possible to link each applicant to a job (a real one or a dummy) and each job to an applicant. That can be done by means of an algorithm that maximizes the sum of the payoff values identified by linking a person to a job. This classifies the applicants and also identifies the ones who are selected versus the ones who are rejected.

2. How to assess the quality of a batch classification model?

Any organization considering or using a batch classification system will undoubtedly want to assess its quality. But how should we express this quality? To begin with, it is important to note that the outcome of such a classification model depends on quite a number of aspects. Let us briefly review some of them.

The outcome is related to the applicant group. The selection ratio together with the level and distribution of relevant aptitudes and characteristics in the group is

obviously of paramount importance.

The outcome is also related to the vacant jobs. These do not only affect the selection ratio but also have a certain level of differentiation as to their attractiveness and the level and profile of aptitudes and characteristics they require. In general, the more differentiated the jobs are, the more powerful the effect of the classification algorithm will be.

The outcome is also highly depending on the payoff computation. The quality of the payoffs depends on things such as the measurement quality of the variables used in the model and their differential validity, the judicious setting of the weights and the integration of metric and categorical data and preferences.

Finally, the classification outcome is conditioned by the chosen objective function and the used algorithm.

The complexity inherent to a batch classification system makes it rather inappropriate to summarize its quality by a single overall number. In many cases the practitioner will be better off with a series of indicators each focusing on a specific aspect of the classification quality. Such indicators are indeed available and can be grouped according to the moment at which they can be obtained.

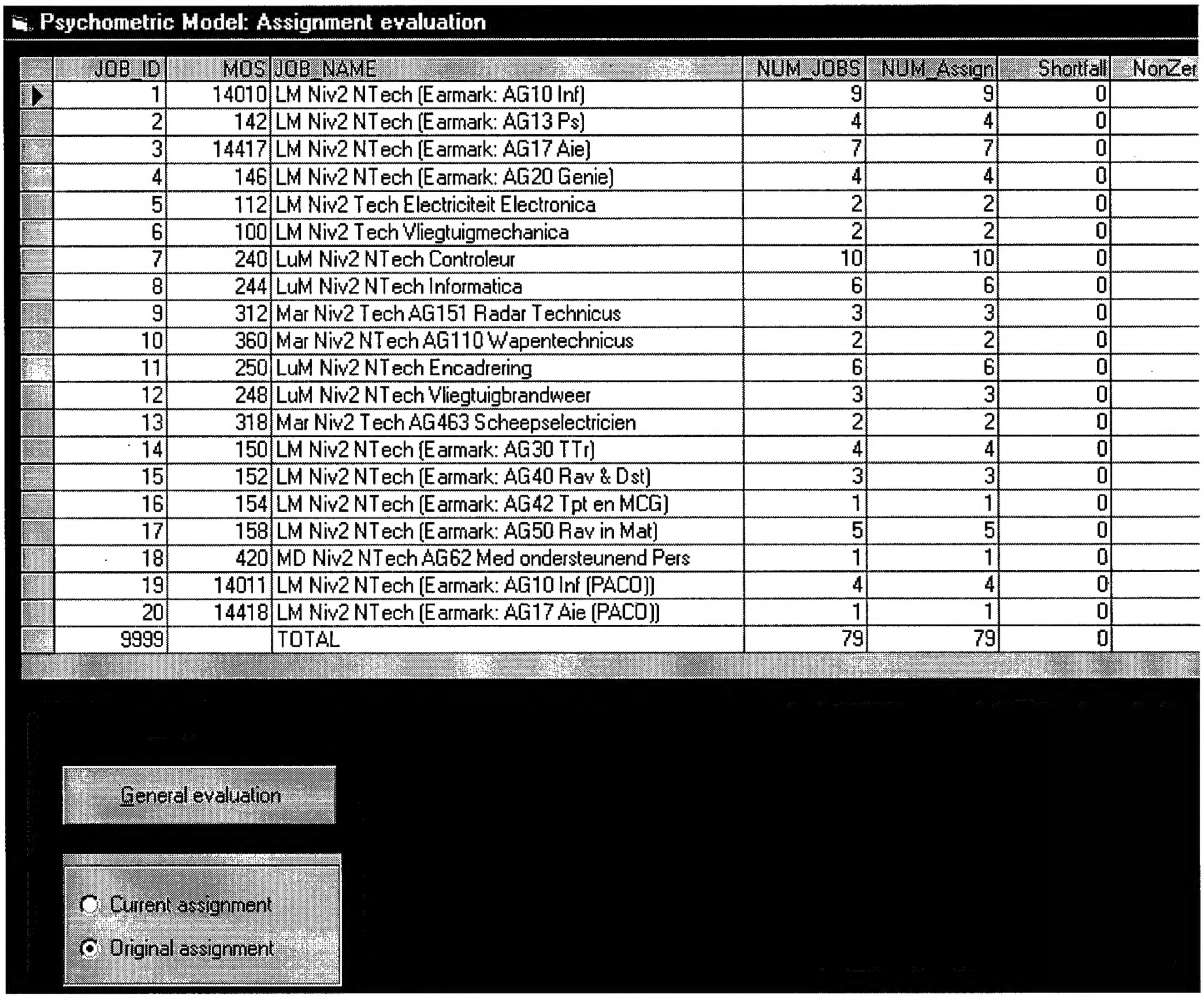
Some indicators depend on data that are not available at the time the classification algorithm is performed. These criterion data typically comprise attrition rates and performance measurements. Quality indicators based on such data include predictive validity coefficients of the payoff-values, differential validity of predictors, logistic regression models against pass-fail criteria, cross checks of the used linear models, etc. Such quality indicators can be called *delayed* or *a posteriori* indicators.

Other quality indicators do not require data which aren't available immediately after the classification algorithm runs. These can be labeled *a priori* or *immediate* quality indicators. Given the title of this paper, we will concentrate our attention on these. These indicators are less powerful than the ones relying on criterion data and cannot provide the practitioner with final statements concerning the quality of the used system, but it offers one tremendous advantage: it allows him or her to modify certain parameters used in the classification model before the assignment decisions are carried out. Put in other words, these indicators allow to detect problems in the classification outcome and to rectify them by altering the parameters of the classification system. The classification model can subsequently be reran until the classification quality is acceptable. It is only at that time that the applicants are informed of the outcome.

We'll now review some immediate quality indicators. To illustrate them, we'll also present some screen views originating from the *Measures of Merit*-module of the *Psychometric Model* which is the batch classification model currently used in the Belgian Armed Forces. The examples come from the classification for the annual Flemish non-commissioned officer recruitment in July, 1998.

2.1. The fill rate.

The first indicator is the fill rate. An important issue is whether or not the vacant jobs will be filled. If the classification model doesn't find suitable applicants for all jobs, how many and which jobs are then left vacant? Did the algorithm have a lot of choice to fill a certain MOS? Are there applicants who didn't get a job but remain available in the event that another candidate resigns for the job he or she got assigned to? These questions can be answered easily for instance by a table like the one presented in following figure.



JOB_ID	MOS	JOB_NAME	NUM_JOBS	NUM_Assign	Shortfall	NonZer
1	14010	LM Niv2 NTech (Earmark: AG10 Inf)	9	9	0	
2	142	LM Niv2 NTech (Earmark: AG13 Ps)	4	4	0	
3	14417	LM Niv2 NTech (Earmark: AG17 Aie)	7	7	0	
4	146	LM Niv2 NTech (Earmark: AG20 Genie)	4	4	0	
5	112	LM Niv2 Tech Electriciteit Electronica	2	2	0	
6	100	LM Niv2 Tech Vliegtuigmechanica	2	2	0	
7	240	LuM Niv2 NTech Controleur	10	10	0	
8	244	LuM Niv2 NTech Informatica	6	6	0	
9	312	Mar Niv2 Tech AG151 Radar Technicus	3	3	0	
10	360	Mar Niv2 NTech AG110 Wapentechnicus	2	2	0	
11	250	LuM Niv2 NTech Encadrering	6	6	0	
12	248	LuM Niv2 NTech Vliegtuigbrandweer	3	3	0	
13	318	Mar Niv2 Tech AG463 Scheepselectricien	2	2	0	
14	150	LM Niv2 NTech (Earmark: AG30 TTr)	4	4	0	
15	152	LM Niv2 NTech (Earmark: AG40 Rav & Dst)	3	3	0	
16	154	LM Niv2 NTech (Earmark: AG42 Tpt en MCG)	1	1	0	
17	158	LM Niv2 NTech (Earmark: AG50 Rav in Mat)	5	5	0	
18	420	MD Niv2 NTech AG62 Med ondersteunend Pers	1	1	0	
19	14011	LM Niv2 NTech (Earmark: AG10 Inf (PACO))	4	4	0	
20	14418	LM Niv2 NTech (Earmark: AG17 Aie (PACO))	1	1	0	
9999		TOTAL	79	79	0	

The first three columns in this table identify the jobs. The next ones give the number of vacant jobs (NUM_JOBS) and the number of persons assigned to them (NUM_Assign). The column 'Shortfall' indicates the number of positions which couldn't be filled. The last two columns give the number of applicants that was eligible for the job (that is, who met all criteria and therefore got a payoff-value for that job) and the number of still 'Available' applicants after the assignment. Those are the ones that have an acceptable payoff but weren't selected in the first place. If the user wants to remedy a shortfall, he or she can lower some thresholds that reject a large number of applicants for that trade or artificially increase the payoffs for the trade so that the algorithm will direct the applicants preferentially to it. A large number of 'available' persons on the other hand, offers the possibility to increase certain minimum thresholds when that is believed to be desirable. One should note

however that usually there is a lot of overlap in the groups of 'available' persons for different trades.

2.2. The Mean Predicted Performance.

The second quality indicator is the Mean Predicted Performance (MPP). Given that the payoff-values are computed using a model based on the relationship between predictors and performance (such as the multiple linear regression model), it becomes possible to estimate the later performance of an individual in a specific trade. After the classification model ran, one can compute the MPP for each trade and compare those with known average performance in the same trades. This quality indicator requires stable prediction models and those are not always available. Its diagnostical power tends to be low as well.

2.3. Descriptive statistics for the groups assigned to trades.

Another approach of the classification quality is based on the descriptive statistics of the groups of applicants that are assigned to the different trades. Aptitudes and other characteristics measured at the interval or ratio level can be summarized by their average whereas categorical data can be shown in contingency tables.

Psychometric Model: Evaluation of means

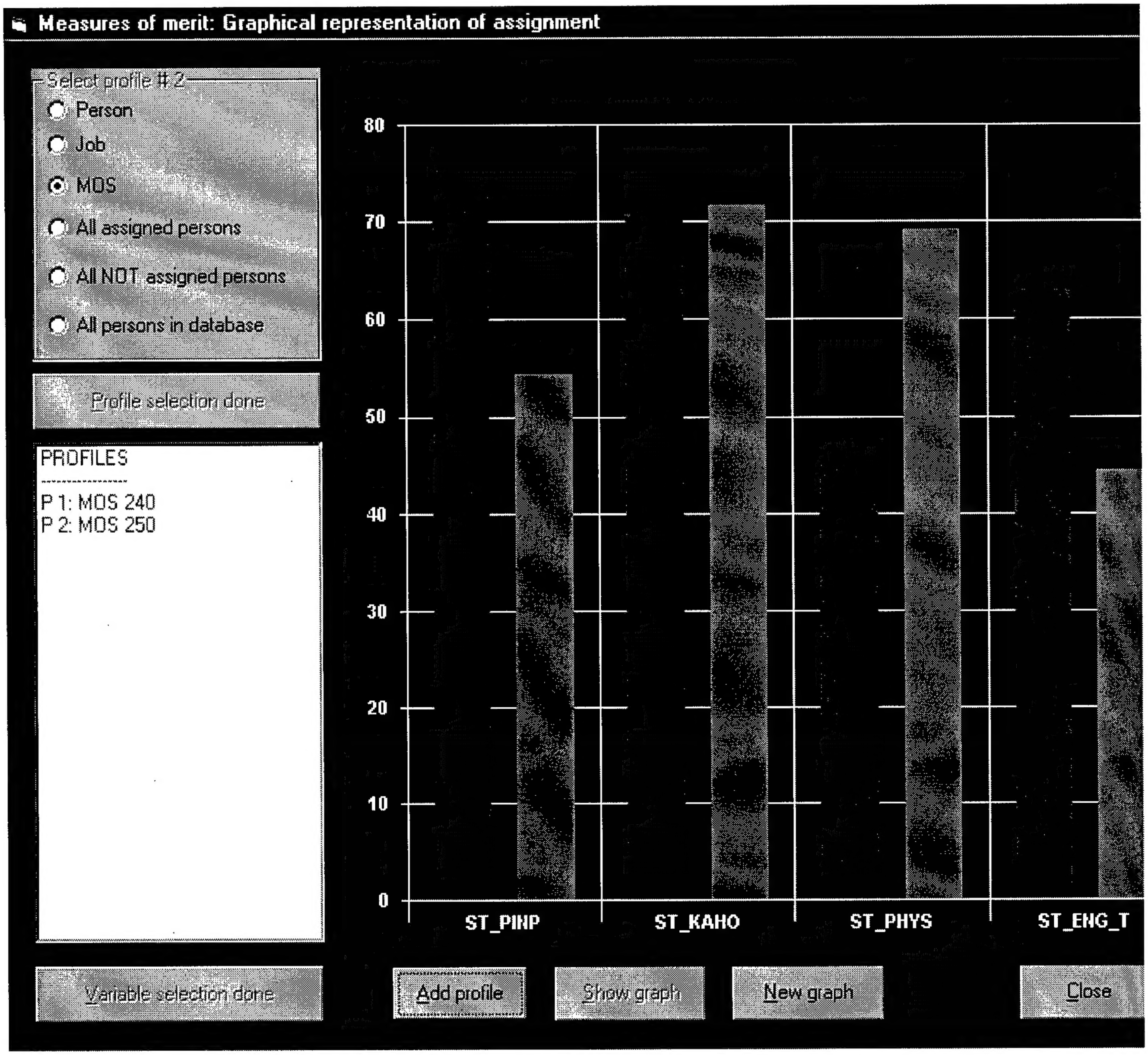
			Mean values for respective groups					
DATANAME	MIN VAL	MAX VAL	ALL	ALL_ASS	ALL_NOT	1	2	3
BENNETB	0	75	47.00273	52.56962	45.47038	52.77778	55.75	56
TRTB	13	73	42.28688	47.10126	40.96167	45.88889	45.75	47
ELECB	18	48	35.20219	36.87342	34.74216	38.33333	39.25	37.85714
PINP	5	495	293.0219	347.076	278.1429	350.3333	368	342
SC_FINAL	4	9	6.508197	7.56962	6.216028	7.444445	7.5	7.857143
PHYSICAL	40	80	59.06557	63.32911	57.89199	66.33334	68	64.85714
ST_PINP	1	99	49.87951	68.40152	44.78112	70.00747	75.94647	65.17239
ST_KAHO	1	99	50	63.88256	46.17867	62.24537	62.97198	67.64313
ST_PHYS	1	99	50	60.06989	47.22815	67.16544	71.10188	63.67888
ELEC	1	99	50.03147	59.94499	47.30266	68.63251	74.08731	65.79884
MECH	1	99	49.99762	64.97481	45.87497	63.37966	67.02995	69.3857
TECH	1	99	50.0089	63.2982	46.35087	65.13061	69.3824	68.19008
ST_AGE	1	99	50	56.7925	48.13029	51.74849	66.90794	52.98599
ST_BENNB	1	99	49.96554	64.3658	46.0017	65.06144	72.81306	73.46506
ST_TRTB	1	99	50.02969	65.58382	45.74825	61.69789	61.24683	65.30634
ATC	0	9	4.879781	5.113924	4.815331	5	5	5
ST_MECH	1	99	49.93992	66.51888	45.37638	64.96995	69.05337	71.68866
ST_ELEC	1	99	50.03147	59.94499	47.30266	68.63251	74.08731	65.79884
ST_TECH	1	99	50.01078	66.46719	45.48096	68.73779	74.00632	72.52887
V1C	1	5	2.103825	2.113924	2.101045	1.333333	1.75	2.285714
V1G	1	5	2.106557	2.35443	2.038327	2.222222	2.25	2.428571
V1H	1	5	1.983607	2.113924	1.947735	1.666667	1.75	2.428571
ENG_G	0	40	33.25956	35.08861	32.7561	34	36.75	35.57143
ST_ENG_G	1	99	50.35348	57.27826	48.44735	52.9464	63.88938	59.19953
ENG_T	0	10	7.68306	8.544304	7.445993	8.222222	9.75	8.571428
ST_ENG_T	1	99	48.89517	56.38197	46.83434	53.19724	68.30386	56.65017
ST_ATC	1	99	51.49006	55.85938	50.28735	53.30899	53.30899	53.30899

Close Refresh Print

The three columns on the left side present the name of the variable and its theoretical minimum and maximum values. The next three columns show the averages for the variable in the row for all applicants in the model (ALL), all assigned applicants (ALL_ASS) and all applicants that were not assigned to a job (ALL_NOT). The

remaining columns show the average of the row-variable for the applicants assigned to the jobs identified by the column-header. When examining the variable ST_PINP for instance (standardized intelligence measurement), one can see that the group of assigned persons has an average of 68.4 whereas the not-assigned group has only 44.7. The persons assigned to the job '2' even have an average of 75.9.

This table is very useful to compare the assigned group versus the not-assigned group to see the selection-effect of the classification model on each variable. This table also contains the necessary data to compare the averaged aptitude profiles for different groups. Such a table however is not very user-friendly for that purpose. That is the reason why another - graphical – instrument was developed. Next figure presents it.



This screen allows to generate graphs very easily. The user can choose any metric variables he or she wants and then select certain profiles. These profiles can include any individual applicant, groups assigned to a specified Job-ID or MOS and the three reference groups: all assigned, all not-assigned and all applicants in the model.

In this example, some average aptitudes are compared for the groups assigned to the MOS *Air Traffic Control* (Profile 1, MOS 240) and *Airfield Defense* (Profile 2, MOS 250). On average, the Air Traffic Control group performs better in General Intelligence (ST_PINP) and Technical English (ST_ENG_T) and lower on Physical Fitness (ST_PHYS). The personality score (ST_KAHO) of both groups is similar. Since this is in accordance with what was desired, no corrective action is required.

Both previous screens focused on metric data. For categorical data, one can check the frequencies of the different variable-classes for several relevant groups.

Psychometric Model: Evaluation of frequencies

DATANAME	CAT	ALL	ALL_ASS	ALL_NOT	1	2	3
FAC_O	1	23	9	14	1	1	1
FAC_O	2	4	2	2	0	0	0
FAC_O	3	0	0	0	0	0	0
FAC_O	4	35	9	26	1	0	1
FAC_O	5	0	0	0	0	0	0
FAC_P	1	0	0	0	0	0	0
FAC_P	2	305	77	228	9	4	7
FAC_P	3	61	2	59	0	0	0
FAC_S	1	0	0	0	0	0	0
FAC_S	2	366	79	287	9	4	7
FAC_S	3	0	0	0	0	0	0
FAC_V	1	197	44	153	3	3	3
FAC_V	2	153	35	118	6	1	4
FAC_V	3	16	0	16	0	0	0
FAC_Y	0	304	59	245	7	3	5
FAC_Y	1	14	7	7	1	1	0
FAC_Y	2	12	4	8	0	0	1
FAC_Y	3	1	0	1	0	0	0
FAC_Y	4	35	9	26	1	0	1
FAC_Y	5	0	0	0	0	0	0
LANGUAGE	1	366	79	287	9	4	7

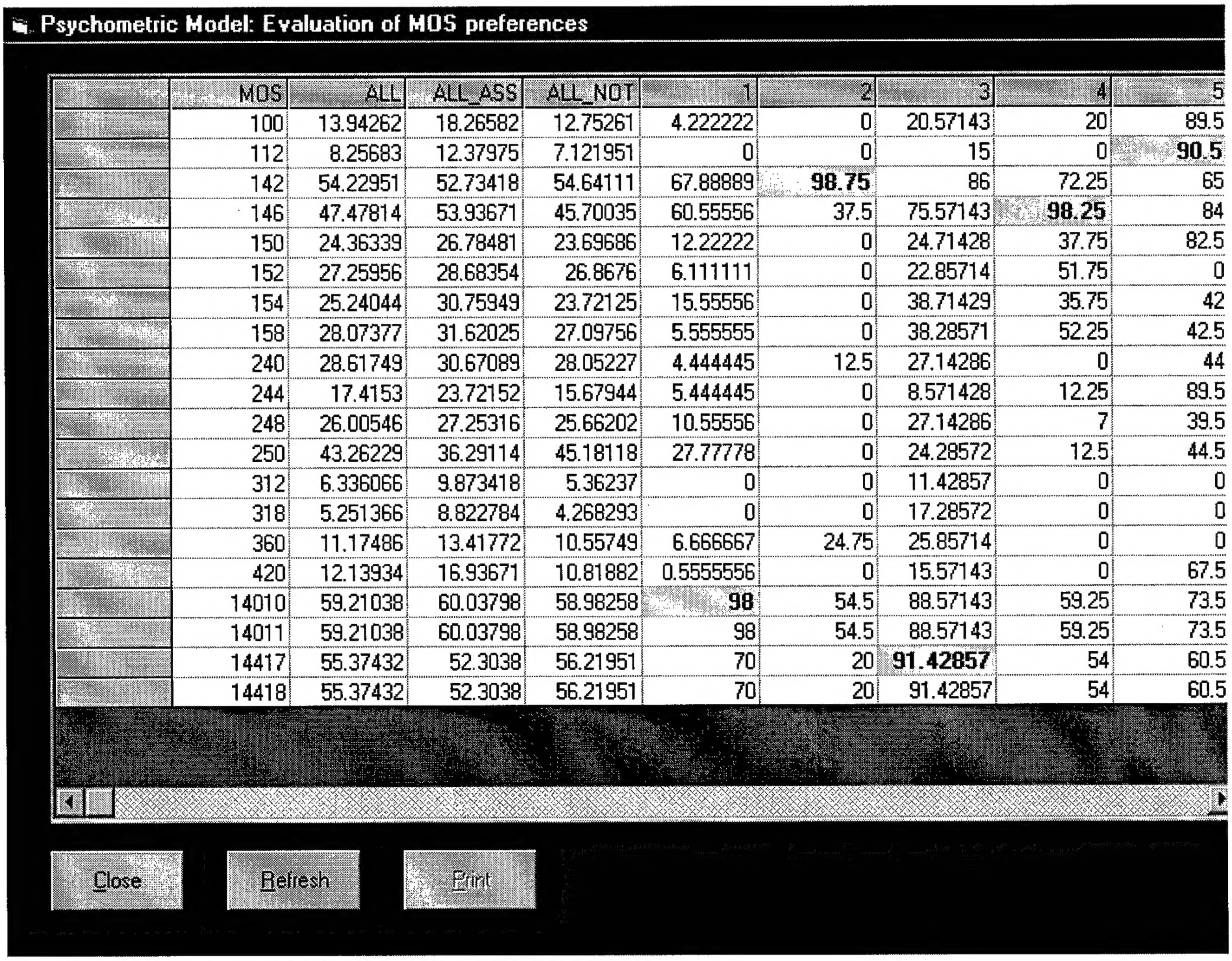
Close Refresh Print

The left column in this table exhibits the categorical variable name and the second column shows the different categories or classes of that variable. The remaining columns contain the observed frequencies of the variable-class in the row for different groups: the three reference groups and the groups assigned to the jobs in the column header.

When looking at the variable 'FAC_P' for instance, which describes the general medical fitness with three classes (1-2-3) that do not exclude the candidate, we notice that no applicant got a FAC_P of 1, 305 applicants got a 2 and 61 of them got a 3. When we look further and use some elementary statistics we can say that the odds to be assigned rather than not-assigned are at least 2.5 times higher for the FAC_P 2 candidates than for the FAC_P 3 applicants (lower bound of 95% exact confidence interval). This can be related to the used coefficients for the classes of the variable to check whether the outcome is desirable.

2.4. Respect of the applicant's preferences.

A modern classification system shouldn't be based on aptitudes only but needs to include the expressed preferences of the applicants as well. When this is the case, it will be of interest to see to what extend the classification model respected the preferences of the applicants. In the *Psychometric Model*, the applicants are requested to express their preferences towards each trade on a 1 to 99 scale. As a quality indicator for the classification model, we'll compute the average preference for a specific trade from the group of applicants that is assigned to that same specific trade.



MOS	ALL	ALL_ASS	ALL_NOT	1	2	3	4	5
100	13.94262	18.26582	12.75261	4.222222	0	20.57143	20	89.5
112	8.25683	12.37975	7.121951	0	0	15	0	**90.5**
142	54.22951	52.73418	54.64111	67.88889	**98.75**	86	72.25	65
146	47.47814	53.93671	45.70035	60.55556	37.5	75.57143	**98.25**	84
150	24.36339	26.78481	23.69686	12.22222	0	24.71428	37.75	82.5
152	27.25956	28.68354	26.8676	6.111111	0	22.85714	51.75	0
154	25.24044	30.75949	23.72125	15.55556	0	38.71429	35.75	42
158	28.07377	31.62025	27.09756	5.555555	0	38.28571	52.25	42.5
240	28.61749	30.67089	28.05227	4.444445	12.5	27.14286	0	44
244	17.4153	23.72152	15.67944	5.444445	0	8.571428	12.25	89.5
248	26.00546	27.25316	25.66202	10.55556	0	27.14286	7	39.5
250	43.26229	36.29114	45.18118	27.77778	0	24.28572	12.5	44.5
312	6.336066	9.873418	5.36237	0	0	11.42857	0	0
318	5.251366	8.822784	4.268293	0	0	17.28572	0	0
360	11.17486	13.41772	10.55749	6.666667	24.75	25.85714	0	0
420	12.13934	16.93671	10.81882	0.5555556	0	15.57143	0	67.5
14010	59.21038	60.03798	58.98258	**98**	54.5	88.57143	59.25	73.5
14011	59.21038	60.03798	58.98258	98	54.5	88.57143	59.25	73.5
14417	55.37432	52.3038	56.21951	70	20	**91.42857**	54	60.5
14418	55.37432	52.3038	56.21951	70	20	91.42857	54	60.5

In this table, the cell values represent the average preference of the group indicated in the column header, for the MOS in the left column. The column 'ALL' indicates the popularity of a MOS. The most relevant cells are highlighted. They represent the preference for a MOS as expressed by the group assigned to that MOS. Low values indicate to a certain extend that the applicants assigned to that trade didn't really want this trade. Very high values could result from giving too much weight to the choices of the applicants, perhaps at the expense of not taking their aptitudes enough into consideration. Problems discovered through this table can be corrected by adapting the weight given to the preferences of the applicants.

2.5. Respect of set profiles.

The following quality indicator attempts to check whether the profiles defined by the weights used to compute the payoffs for a trade, correspond to the aptitude profiles of the applicants assigned to that trade. To do so, one needs to consider the variables used to calculate the payoffs for a specific trade. If you standardize these over all the acceptable applicants to a common mean and variance, and then take the average on these standardized variables for the group of applicants assigned to that trade, one can see the departure from the overall mean as an indicator of the weight *actually* given to the variable in the model. It is further possible to express these trade-averages and the weights used to compute the payoffs on the same scale and to compare them pairwise. This can be done graphically or by means of correlations.

2.6. Specificity of set profiles.

The last proposed quality indicator consists of the correlation matrix of the payoffs. Highly positively correlated payoffs indicate a possible lack of differentiation between the requested aptitude profiles. If the concerned trades

are not considered to be very similar, one should try to identify means to discriminate between them and to incorporate these in the classification model.

3. Future directions

When using the immediate quality indicators as described, a practitioner can get a very accurate idea of the quality of the used batch classification system. Such a quality assessment however, still requires a good amount of expertise. Therefore it is recommended to develop expert systems detecting problems and suggesting ways to correct them to assist the user of such classification systems.

References

DARBY, M., SKINNER, J. and ALLEY, W (1996)*A Methodology for Evaluating the Classification Potential of Experimental Tests.*. Proceedings of the 38th annual conference of the International Military Testing Association, p 285 - 289.

DARBY, M., ALLEY, W and CHENG, C. (1996) *The Practical Benefits of Personnel Testing: An extension of the Taylor-Russell Tables to Multiple Job Categories*. Ibidem p. 268-273.

DARBY, M., GROBMAN, J. et al. (1996) *The Generic Assignment Test and Evaluation Simulator (GATES), user manual*. Human Resources Directorate, Manpower and Personnel Research Division, USAF.

LESCREVE, F. (1993) *A Psychometric Model for Selection and Assignment of Belgian NCO's*. Proceedings of the 35th Military Testing Association. US Coast Guard, p. 527-533.

LESCREVE, F. (1995) *The Selection of Belgian NCO's: The Psychometric Model goes operational*. Proceedings of the 37th annual conference of the International Military Testing Association. Canadian Forces Personnel Applied Research Unit, p. 497-502.

LESCREVE, F. (1997) *Data modeling and processing for batch classification systems*. Proceedings of the 39th annual conference of the International Military Testing Association (in press)

INTERNET DOCUMENT INFORMATION FORM

A . Report Title: Immediate Assessment of Batch Classification Quality

B. DATE Report Downloaded From the Internet 4/21/99

C. Report's Point of Contact: (Name, Organization, Address, Office Symbol, & Ph #): Navy Education and Training Professional Development and Technology Center
Navy Advancement Center Dept
Dr. Grover Diesel, (850) 452-1815
6490 Saufley Field Road
Pensacola, FL 32509-5237

D. Currently Applicable Classification Level: Unclassified

E. Distribution Statement A: Approved for Public Release

F. The foregoing information was compiled and provided by:
DTIC-OCA, Initials: VM_ **Preparation Date**: 4/22/99__

The foregoing information should exactly correspond to the Title, Report Number, and the Date on the accompanying report document. If there are mismatches, or other questions, contact the above OCA Representative for resolution.